



Regresión Lineal

Rodolfo Torres 1742143
ERICK ULRICH



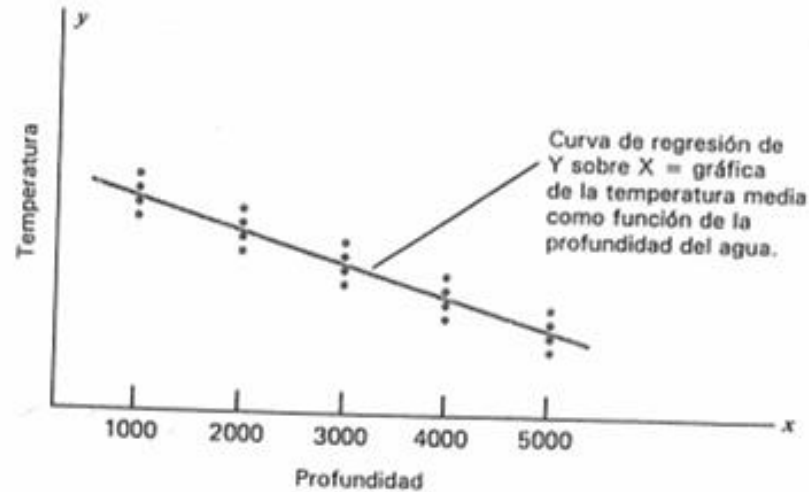
¿Que es la regresión Lineal?

En la regresión buscamos una variable aleatoria simple digamos Y, En teoría el valor de esta variable aleatoria está influenciado por los valores tomados por una o más variables.

Y se denomina como: "Variable Dependiente" o "Respuesta"

Las variables influyentes: "Variables Independientes" , "Variable Predictoras" o regresoras

Al realizar una predicción los regresores no se tratan como variables al azar , son entidades que asumen valores diferentes, pero no es al azar.



Superficie m*m	Alquiler(\$)
150	450
120	380
170	480
80	270

- Queremos modelar el precio de los pisos en una ciudad según su superficie en metros cuadrados.
- La entrada o variable predictora nos ayuda a predecir el valor de la variable de salida(superficie).
- La variable de salida es la que queremos predecir.

- En el caso de la regresión lineal asumimos que Y(costo) es una función lineal de x (superficie) y entonces el modelo lineal se escribe como

$$Y_e = \alpha + \beta * x$$

$$\text{Alquiler mensual} = \alpha + \beta * \text{Superficie}$$

Con datos históricos, podríamos crear un modelo lineal y obtener las posibles valores de alpha y Beta.

$$\alpha = 86.96, \beta = 2.37$$

Modelo lineal

$$Y_e = 86.96 + 2.37 * x$$



Usando esta ecuación, podremos encontrar el alquiler de cualquier casa, por ejemplo una de 110 m²:

$$Y_e = 86.96 + 2.37 * 110 = 347.66$$

Como podemos encontrar los parámetros alfa y beta.

La diferencia entre el valor real y el estimado se puede escribir como:

$$e_i = (Y_i - Y_e(X_i))$$

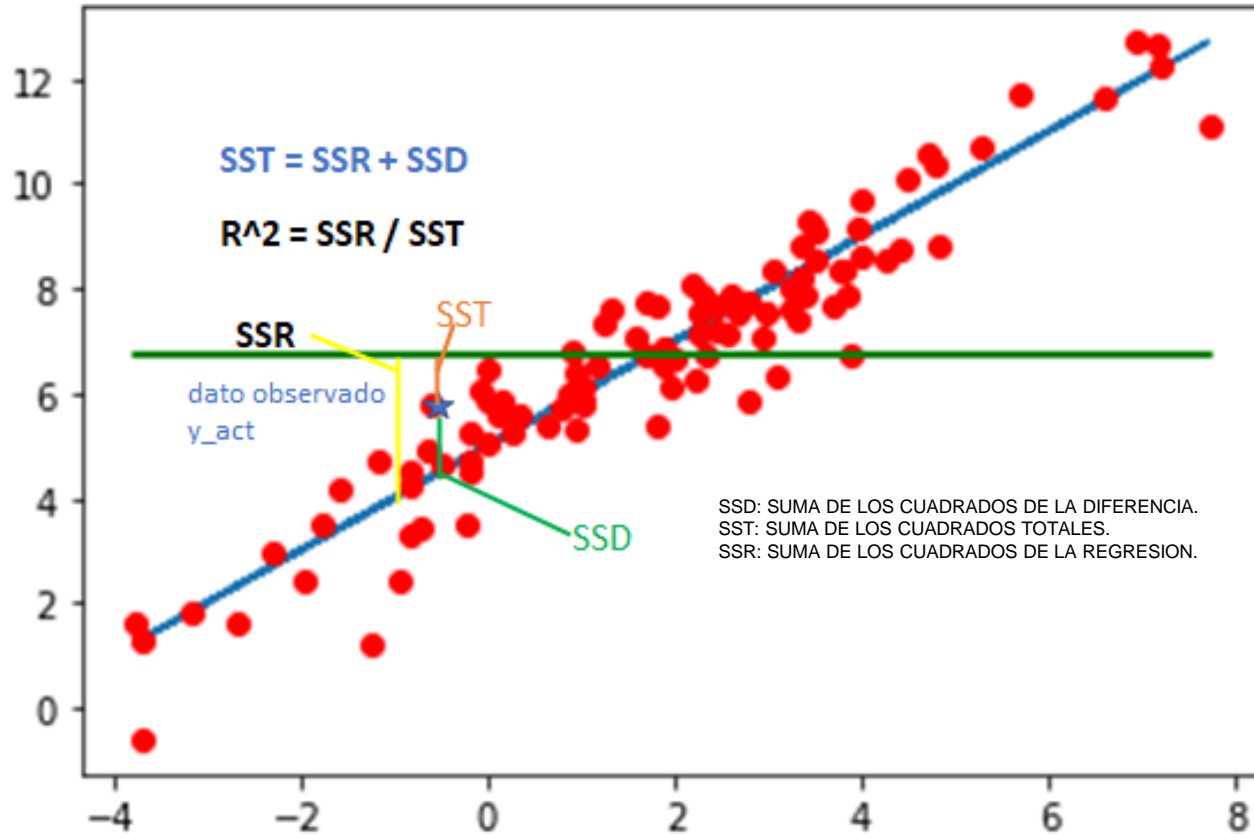
El objetivo es encontrar los valores de α y β que minimicen la suma de los cuadrados de las diferencias entre los valores reales y los estimados para todos los puntos del conjunto de datos.

$$\min \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - Y_e(X_i))^2 = \sum_{i=1}^n (Y_i - (\alpha + \beta * X_i))^2$$

$$\beta = \frac{\sum_{i=1}^n (X_i - x_m)(Y_i - y_m)}{\sum_{i=1}^n (X_i - x_m)^2}$$

$$\alpha = y_m - \beta * x_m$$

Valor Actual vs Prediccion



Componente del error

El modelo en un mundo ideal sería perfectamente lineal

$$Y_e = \alpha + \beta * x$$

Realmente siempre tendrá un componente de error o residuo E

$$Y_e = \alpha + \beta * x + E$$

El residuo E sera una variable aleatoria con distribución normal.

El p valor

El modelo presenta una relación lineal entre x e y

$$f(x) = \begin{cases} H_0 : \beta = 0 \\ H_0 : \beta \neq 0 \end{cases}$$

Para comprobar la existencia de una relación lineal entre x e y, rechazamos el contraste de hipótesis

Si el p-valor resultante es menor que el nivel de significancia, rechazamos la hipótesis nula y aceptamos que existe una relación lineal entre x e y.

Error estándar residual

- RSE es la desviación estándar del termino del error(desviación de la parte de daos que el modelo no es capaz de explicar por falta de información o mas datos adicionados)
- En el caso de un regresión lineal simple

- $$RSE = \sqrt{\frac{\sum(Y_i - Y(X_i))^2}{n-2}} = \sqrt{\frac{SSD}{n-2}}$$

- En el caso de una regresión lineal múltiple (K= numero de variables predictivas)

- $$RSE = \sqrt{\frac{\sum(Y_i - Y(X_i))^2}{n-k-2}} = \sqrt{\frac{SSD}{n-k-2}}$$